WiNLP 2025

The 9th Widening NLP Workshop

Proceedings of the Workshop

November 8, 2025

The WiNLP organizers gratefully acknowledge the support from the following sponsors.

Platinum

©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-351-7

Introduction

Welcome to the 2025 Widening NLP Workshop!

The origins of this workshop trace back to ACL 2016 in Berlin, where a small group gathered to address the underrepresentation of women and other minorities in the Natural Language Processing (NLP) community. That conversation led to the inaugural Workshop for Women and Underrepresented Minorities in NLP at ACL 2017—a dedicated space to highlight the voices and contributions that are too often overlooked. Since then, Widening NLP has continued to grow, and we are proud to carry this important tradition forward in 2025.

Over the years, we have taken deliberate steps to make the workshop more inclusive. Following the inaugural 2017 event, we introduced two submission deadlines—an early one to support those requiring additional time for visa applications and a later one for others without such constraints. Building on the success of 2018, the 2019 edition expanded our focus to celebrate diversity not only in gender and minority identity but also in scientific background, discipline, degree, training, and seniority. That year, we also launched a peer feedback system, giving authors the chance to receive constructive input from colleagues prior to formal review.

The global pandemic brought new challenges by canceling in-person events, yet it also created opportunities to broaden access. Since 2021, we have embraced a hybrid format that enables us to welcome a wider audience while continuing to support in-person participation. In addition, we repurposed travel funds to cover high-speed internet and registration costs, ensuring that even more participants could engage in the virtual workshop.

This year, we are excited to present a truly outstanding program. From 76 submissions, 41 papers were accepted, reflecting both the quality and diversity of perspectives within our community. We are honored to feature two distinguished invited speakers - Jen-Tse Huang and David Adelani, as well as four excellent panelists - Christos Christodoulopoulos, Julia Kreutzer, Pittawat Taveekitworachai, Zhisong Zhang. Their talks and discussions will inspire us with the wealth of talent and ideas driving the advancement of NLP today.

We warmly welcome you to the 2025 Widening NLP Workshop. We hope you will find inspiration in the work of our authors, speakers, and panelists, and that this gathering will continue to foster new connections, collaborations, and possibilities for an ever more inclusive NLP community.

-Chen, Emily, Hua, Lesly, Yinqiao, Meryem, Peerat, Richard, Santosh, Sophia, Surendrabikram, Wiem, Organizing Co-Chairs

Organizing Committee

Organizing Chairs

Chen Zhang, Peking University
Emily Allaway, University of Edinburgh
Hua Shen, New York University (Shanghai) / University of Washington
Lesly Miculicich, Google
Yinqiao Li, City University of Hong Kong
Meryem M'hamdi, Meta
Peerat Limkonchotiwat, AI Singapore
Richard He Bai, Apple
Santosh T.Y.S.S., Amazon
Sophia Simeng Han, Yale University
Surendrabikram Thapa, Virginia Tech, USA
Wiem Ben Rim, University College London

Advisory Board

Alham Fikri Aji, MBZUAI & Google Research
Helena Gomez-Adorno, IIMAS, UNAM
Sarvnaz Karimi, CSIRO
Sunayana Sitaram, Microsoft Research India
Viviane Moreira, UFRGS - Brazil
Zeerak Talat, University of Edinburgh

Program Committee

Program Chairs

Emily Allaway, University of Edinburgh
Yinqiao Li, City University of Hong Kong
Meryem M'hamdi
Santosh T.y.s.s, Amazon
Chen Zhang, Peking University

Area Chairs

Emily Allaway, University of Edinburgh
Richard He Bai, Apple
Simeng Han, Yale University
Peerat Limkonchotiwat, AI Singapore
Meryem M'hamdi
Lesly Miculicich, Google
Chen Zhang, Peking University

Reviewers

Amir Abdullah, Giuseppe Abrami, David Alfter, Evelin Amorim, Raviteja Anantha, Arturo Argueta, Akshatha Arodi, Ekaterina Artemova

Nikolay Babakov, JinYeong Bak, Yuwei Bao, Leslie Barrett, Dario Bertero, Aditya Bhargava, Kasturi Bhattacharjee, Mukul Bhutani, Emanuela Boros, Daniel Braun

Sky CH-Wang, Rémi Cardon, Yekun Chai, Sunandan Chakraborty, Andong Chen, Bo Chen, Guanyi Chen, Yue Chen, Ziyang Chen, Elena Chistova, Won Ik Cho, Juhwan Choi, Seungtaek Choi

Wen Dai, Debarati Das, Brian Davis, Prajit Dhar, Wentao Ding, Nemanja Djuric, Phong Nguyen-Thuan Do, Xiangjue Dong, Nisansa de Silva

Micha Elsner

Neele Falk, Ge Fan, Shangbin Feng, Alejandro Figueroa, Margaret M. Fleck, Shuai Fu

Baban Gain, Prakhar Ganesh, Harritxu Gete, Sourav Ghosh, Sreyan Ghosh, Sucheta Ghosh, Hyojun Go, Anmol Goel, Venkata S Govindarajan, Navita Goyal, Loïc Grobol, Varun Gumma, Tunga Gungor, Jialiang Guo, Zhen Guo

Yo-Sub Han, Peter Hase, Estrid He, Yu Hou, Minghui Huang, Ben Hutchinson

Mert Inan

Rishabh Jain, Sébastien Jean, Jiyue Jiang, Nan Jiang, Kenneth Joseph

Kazuma Kadowaki, Pride Kavumba, Byoungjip Kim, Gyuwan Kim, YoungBin Kim, Tracy Hol-

loway King, Svetla Peneva Koeva, Michalis Korakakis, Katsunori Kotani, Elisa Kreiss, Ralf Krestel, Satyapriya Krishna, Shivani Kumar, Kemal Kurniawan, Mascha Kurpicz-Briki

Joosung Lee, Chong Li, Dongyuan Li, Ruifan Li, Ruosen Li, Yinqiao Li, Jasy Suet Yan Liew, Gilbert Lim, Peerat Limkonchotiwat, Haowei Lin, Lucy H. Lin, Alisa Liu, Danni Liu, Fuxiao Liu, Yujie Lu, Evan Lucas, Li Lucy, Jixiang Luo, Yiran Lawrence Luo, Zhekun Luo, Pedro Henrique Luz de Araujo

Ziqiao Ma, Fred Mailhot, Magdalena Markowska, Marcos Martínez Galindo, John Philip McCrae, Alexander Mehler, Lesly Miculicich, Simon Mille, Hideya Mino, Ashutosh Modi, Anjishnu Mukherjee, Sheshera Mysore

Diane Napolitano, Youyang Ng, Kiem-Hieu Nguyen, Sergiu Nisioi, Tadashi Nomoto, Enrique Noriega-Atala, Damien Nouvel

Eda Okur, Naoki Otani

Aline Paes, Letitia Parcalabescu, ChaeHun Park, Kunwoo Park, Alicia Parrish, Yifan Peng, Van-Thuy Phi, Fred Philippy, Aidan Pine, Rajesh Piryani, Sukannya Purkayastha, Rifki Afina Putri

Leonardo Ranaldi, Priya Rani, Hannah Rashkin, Rezvaneh Rezapour, Matīss Rikters, Brian Roark, Angelika Romanou, Susanna Rücker

Fatiha Sadat, Brenda Salenave Santana, Anastasiia Sedova, Sofia Serrano, Rita Sevastjanova, Hua Shen, Qinlan Shen, Quan Z. Sheng, Ning Shi, Kazutoshi Shinoda, Yow-Ting Shiue, Mei Si, Li Siyan, Hyun-Je Song, Katherine Stasaski, Elias Stengel-Eskin

Santosh T.y.s.s, Eric S. Tellez, Hrishikesh Terdalkar, Dimitrios Tsarapatsanis

Can Udomcharoenchaikit, Stefan Ultes, David Uthus

Sowmya Vajjala

Hai Wang, Jiaan Wang, Qingyun Wang, Ruibo Wang, Yanhao Wang, Leonie Weissweiler

Kaige Xie, Bo Xu, Jinan Xu, Wenduan Xu

Shuntaro Yada, Changbing Yang, Li Yang, Ken Yano, Yuwei Yin, Hiyori Yoshikawa, Mengxia Yu

Qingcheng Zeng, Chen Zhang, Ningyu Zhang, Wei Emma Zhang, Xiang Zhang, Zecheng Zhang, Zequn Zhang, Mengjie Zhao, Yizhou Zhao, Zheng Zhao, Zhenjie Zhao, Ji-Zhe Zhou, Henghui Zhu, Heike Zinsmeister

Keynote Talk

Language Models Do Not Have Human-Like Working Memory

Jen-Tse Huang
Johns Hopkins University

Abstract: While Large Language Models (LLMs) exhibit remarkable reasoning abilities, we demonstrate that they lack a fundamental aspect of human cognition: working memory. Human working memory is an active cognitive system that enables not only the temporary storage of information but also its processing and utilization, enabling coherent reasoning and decision-making. Without working memory, individuals may produce unrealistic responses, exhibit self-contradictions, and struggle with tasks that require mental reasoning. Existing evaluations using N-back or context-dependent tasks fall short as they allow LLMs to exploit external context rather than retaining the reasoning process in the latent space. We introduce three novel tasks: (1) Number Guessing, (2) Yes-No Deduction, and (3) Math Magic, designed to isolate internal representation from external context. Across seventeen frontier models spanning four major model families, we consistently observe irrational or contradictory behaviors, indicating LLMs' inability to retain and manipulate latent information. Our work establishes a new benchmark for evaluating working memory in LLMs and highlights this limitation as a key bottleneck for advancing reliable reasoning systems.

Bio: Jen-Tse (Jay) Huang is a postdoctoral researcher at the Center for Language and Speech Processing (CLSP) at Johns Hopkins University, working with Mark Dredze. He received his Ph.D. in Computer Science and Engineering from the Chinese University of Hong Kong and his B.Sc. from Peking University. His research explores the evaluation of large language models (LLMs), both as individual agents and as collectives in multi-agent systems, through the lens of social science. His work has been published in top-tier AI venues, including an oral presentation at ICLR 2024. He actively serves as a reviewer for major conferences and journals such as ICML, NeurIPS, ICLR and serves as an area chair in ARR.

Keynote Talk

Scaling Multilingual Evaluation of LLMs to Many Languages

David Adelani
McGill University

Abstract: Despite the widespread adoption of Large language models (LLMs), their remarkable capabilities remain limited to a few high-resource languages. In this talk, I would describe different approaches to scaling evaluation to several languages. First, I would describe simple strategies for extending multilingual evaluations by re-purposing existing English datasets to over 200 languages for both text (SIB-200) and speech modalities (Fleurs-SLU). Second, I would introduce our recent bench IrokoBench – a human-translated benchmark dataset for 17 typologically-diverse low-resource African languages covering three tasks: natural language inference, mathematical reasoning, and multi-choice knowledge-based question answering. This evaluation expands the evaluation of many low-resource languages from simple text classification tasks to more challenging knowledge and reasoning tasks. We observe a significant performance gap between open and proprietary models, with the highest performing open model, Gemma 2 27B, only at 60% of the best-performing proprietary model GPT-4o performance. These findings suggest that more efforts are needed to develop and adapt LLMs for low-resource languages. Finally, I will highlight some of our recent projects that make some of these challenging datasets more multicultural for Visual question answering and intent detection tasks, to encourage practical usage of LLMs within the low-resource communities.

Bio: Dr. David Adelani is an Assistant Professor at the McGill University School of Computer Science, a Core Academic Member at Mila - Quebec AI Institute, an IVADO Professor, and a Canada CIFAR AI Chair. He received his Ph.D in Computer Science at the Department of Language Science and Technology, Saarland University, Germany. His research interests include multilingual natural language processing with a focus on low-resource languages, speech processing, privacy and safety of large language models. With over 20 publications in leading NLP and Speech Processing venues like ACL, TACL, EMNLP, NAACL, COLING, and Interspeech, he has made significant contributions to NLP for low-resource languages. Notably, one of his publications received the Best Paper Award (Global Challenges) at COLING 2022 for developing AfroXLMR, a multilingual pre-trained language model for African languages. Other notable awards include an Area Chair Award at IJCNLP-AACL 2023, Outstanding Paper Award and Best Theme Paper Award at NAACL 2025.

Panel

After a PhD, What is Waiting for us? A Discussion and Experiences from Industry, Academia, and Startups

Christos Christodoulopoulos, Julia Kreutzer, Pittawat Taveekitworachai, Zhisong Zhang

Bio:

Christos Christodoulopoulos

Christos Christodoulopoulos is a Principal Technology Adviser in the AI Policy & Compliance teams of the Information Commissioner's Office, UK's Data Protection regulator. Before joining the ICO, he was an Applied Scientist at Amazon, starting in 2016 on the Alexa AI Knowledge team and ending as a Senior Applied Scientist at Amazon's Responsible AI team working on multimodal and agentic FM development. Before Amazon, he was a postdoctoral researcher at UIUC working with Dan Roth on Semantic Role Labeling and Cindy Fisher on computational models of child language acquisition. He has an MSc and PhD from the University of Edinburgh. He is a Program Chair for EMNLP 2025, an organiser for the FEVER, GenBench, and TrustNLP workshops and has served as a reviewer, area chair and senior area chair for many *CL conferences.

Julia Kreutzer

Julia Kreutzer is a Senior Research Scientist at Cohere Labs, where she conducts research on large language models, currently focused on multilinguality, evaluation and inference. Previously, she worked at Google Translate, and completed her PhD at Heidelberg University on learning from human feedback in machine translation. She's been an active contributor to multiple open-science communities and a co-organizer of COLM, WMT shared tasks and various NLP workshops.

Pittawat Taveekitworachai

Pittawat (Pete) Taveekitworachai is a research scientist on the Typhoon team at SCB 10X in Thailand. His research interests include reasoning models, test-time scaling, prompt engineering, and reinforcement learning. He completed his Master's degree (as valedictorian) at Ritsumeikan University, Japan, under the Japanese Government Scholarship (MEXT), where his research focused on prompt engineering, large language models, and their applications in gaming, healthcare, and autonomous driving. At SCB 10X, he leads research collaborations with academic and industry partners, both domestically and internationally. He is passionate about translating cutting-edge research into real-world applications and values both the scientific rigor and engineering practicality that drive impactful innovation.

Zhisong Zhang

Zhisong Zhang is currently an Assistant Professor in the Department of Computer Science of City University of Hong Kong. He holds a PhD from the Language Technologies Institute at Carnegie Mellon University. His doctoral research focused on advancing natural language processing (NLP) systems, particularly in data-limited scenarios, where his work aimed to reduce the need for labor-intensive manual data labeling while improving task performance. After PhD graduation, he had also worked as a researcher in Tencent before joining CityUHK. His current research focuses on natural language processing (NLP) and large language models (LLMs), with particular interests in long-context language modeling, LLM-based agent systems, and understanding the underlying mechanisms of language models. Please refer to his homepage for more details: https://zzsfornlp.github.io/

Table of Contents